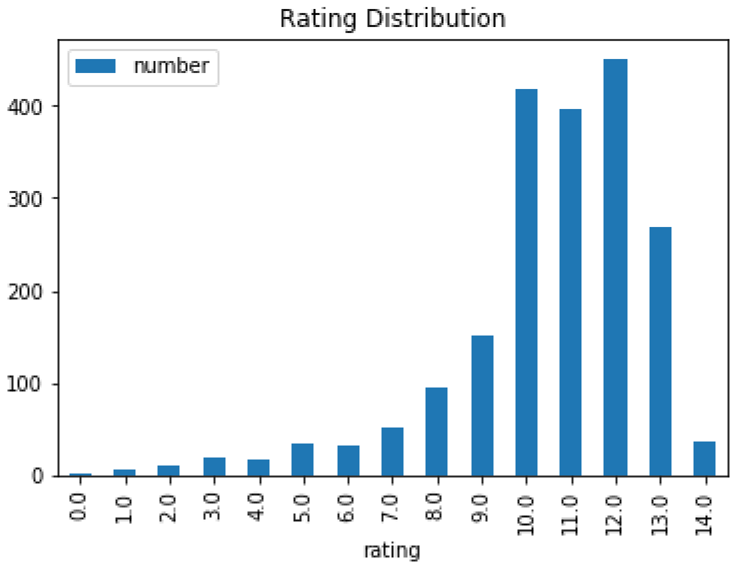
After cleaning all 3 data frames and merging them into one, we can utilize the describe pandas method, which generates a descriptive summary of the data frame object it was called from. It includes statistical descriptions of the data frame such as count of each variable or feature; it includes the mean, standard deviation, minimum value, 25, 50 and 75 percentile as well as the maximum value.

From this table, I was able to answer questions such as what is the most favorite tweet, which happens to be a puppy with a sign that reads "I MARCH 4 MY MOMS", the dog is supposed to be supporting the Women's March. It was rated 13/10

The most retweeted was actually a video and not a photo. It was a video of a dog that was in a pool standing and it would walk in similar fashion as a human would. This dog was rated a 13/10.

In terms of which tweets were the highest rated ones, I did not pay much attention to that since I actually excluded rating numerators with 3 digits since I consider them outliers. They were skewing the rating distribution with ratings referencing dates, years, and other jokes. If it matters, the highest numerator was 1776 in reference to USA Independence Day for a dog dressed with the American flag.

Finally, in our visual titled "Rating Distribution" if we only consider ratings from 0 to 14, we see that the most common rating was 12. This surprised me because from glancing at the data, I would constantly see 13 ratings, and I thought 13 must have been the most repeated. The visual doesn't reflect this, instead 13 is not even in the top 3 most frequent numerators.


Rating Distribution
number
rating